Psychotherapy
2012, Vol. 49, No. 2, 115–122

0033-3204/12/$12.00 DOI: 10.1037/a0027932

The Treatment Outcome Package: Facilitating Practice and Clinically Relevant Research

Soo Jeong Youn
Pennsylvania State University

David R. Kraus
Outcome Referrals, Inc., Marlborough, Massachusetts

Louis G. Castonguay
Pennsylvania State University

As an effort to foster the use of standardized data in clinical practice, this article discusses an outcome measure developed by clinicians for use in naturalistic settings: the treatment outcome package (TOP). In addition to describing how the TOP can be used in day-to-day practice, the article illustrates how its multidimensional structure can inform several aspects of clinical work, such as the development of case formulations, facilitation of client–therapist communication, tracking (positive and negative) change during treatment, and the documentation of specific area of therapist expertise. This article recognizes the challenges and drawbacks that are associated with the use of a standardized outcome measure and also describes three ways by which the TOP can contribute to the strengthening of the relationship between research and practice.

Keywords: outcome measure, naturalistic setting, clinically relevant research, monitoring of therapeutic progress

Before applying to graduate school, students wanting to become therapists can choose many pathways to do so. Yet, of all the professional roads leading to the practice of psychotherapy, one of the most often chosen (not the least competitive and certainly not the most remunerative) is arguably the most rooted in science. Many individuals who apply to doctoral programs in psychology (clinical or counseling) do so, at least in part, because they believe that scientific knowledge about human functioning and therapy can be relevant to the understanding and treatment of psychopathology. As such, they should be excited by data and the opportunity to collect it. However, it is well recognized that most clinicians (including psychologists) are not making use of standardized data as part of their day-to-day practice. As such, they are not formally tracking client progress during therapy, nor are they documenting the potential therapeutic benefits gained by clients at the end of treatment. The goal of this article is to present a standardized measure, the Treatment Outcome Package (TOP; Kraus, Seligman, & Jordan, 2005), that has been developed to make outcome data collection not only friendly for clinicians, but also beneficial for both the clinician and his or her clients. This article recognizes the challenges and drawbacks that can be associated with the use of standardized outcome assessment and describes how the TOP can help practitioners address crucial clinical tasks, such as case formulations and treatment planning, as well as documenting change during treatment. Additionally, the use of an outcome measure such as the TOP can pave a stronger bridge between science and practice by helping clinicians develop research in naturalistic settings, as well as seamlessly integrate clinical and empirical tasks. Ultimately, the use of psychometrically rigorous and clinically relevant outcome can help achieve a goal that is likely to be valued by clinicians of different theoretical orientations and professional backgrounds, improving the services they provide to the wide variety of clients they work with in their everyday practice.

The TOP

Adopted by a large number of mental health providers, the TOP is an outcome measure specifically designed for natural settings (Kraus & Castonguay, 2010). Based on years of exploratory and confirmatory factor analytic work, it has been constructed to provide an extensive assessment of clients' difficulties and resources while being short enough to be repeatedly administered without adding burden to the clients and providers. The TOP questions have high face validity to patients and psychotherapists alike, they are easy to read (fifth-grade level), and are related to DSM symptoms that are key to an initial interview (e.g., "felt little or no interest in most things"). The TOP has three age versions (child, adolescent, and adult), and each of these contain a companion clinician-rating tool. The adult version of the TOP includes 58 items that load onto 12 scales measuring a broad range of issues related to symptomatology and life functioning, including depression, panic, psychosis, suicidal ideation, violence, mania, sleep, substance abuse, social conflict, work functioning, sexual functioning, and quality of life. As a multidimensional measure, it is based on the concept that

Soo Jeong Youn and Louis G. Castonguay, Department of Psychology, Pennsylvania State University; and David R. Kraus, Outcome Referrals, Inc., Marlborough, Massachusetts.

Correspondence concerning this article should be addressed to Soo Jeong Youn, Department of Psychology, Pennsylvania State University, 346 Moore Building, University Park, PA 16802. E-mail: svy5080@psu.com

psychopathology manifests itself in different ways depending on clients' needs and difficulties. Because the scores are standardized, the level of difficulty for all of the issues measured can be compared with each other within the same scale. Thus, the TOP is usually administered by itself, without the need of additional measures. In addition to the 12 subscales, the TOP also measures demographics, health, substance use, stressful life events, treatment goals, and satisfaction with treatment, and diagnostic considerations are reported for Axes I, III, and IV. With a data set of over a million patients, the company that processes the TOP—Behavioral Health Laboratories, Inc. (BHL)—can provide risk-adjusted benchmarking of each clinician's strengths and weaknesses that can help clinicians prioritize the patients they should treat and the continuing education programs with which they may wish to focus their professional development.

The TOP has excellent factorial structure, as found in exploratory and confirmatory factor analyses with five large client samples (Kraus et al., 2005). All reliabilities for subscales were excellent (ranging from .87 to .94), with the exception of the Mania subscale that has a lower but acceptable reliability of .76, which is due to the bimodal distribution of mania items (see Table 1). Excellent level of convergent validity has also been shown with the Beck Depression Inventory (BDI), the Brief Symptom Inventory (BSI), the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), the Basis 32, and the SF-36. The TOP Depression scale, for example, correlates .92 with the BDI (see Kraus et al., 2005, for all correlations). In addition, the TOP has been demonstrated to have virtually no ceiling effects, suggesting that it can capture the severe extremes of the clinical constructs it is aimed to assess. Furthermore, each TOP subscale measures at least a half to more than two standard deviations into the "healthy" tails of its constructs, which suggests that each TOP scale measures a wide range of clinical severity. The TOP has also been found to be sensitive to change: with an average of only seven treatment sessions, Cohen's *d* effect sizes ranged from .27 (Sexual functioning) to .91 (Depression) (see Table 1). Most TOP measures showed reliable improvement for at least a quarter of participants, and 91% of clients showed reliable improvement on at least one TOP subscale. As shown in Table 1 under the column "Reliable change," we present the amount of change in TOP subscale required to document reliable change according to Jacobson and Truax (1991). These reliable change scores range from .48 to .96. Finally, the TOP is able to predict with some degree of accuracy whether an individual is a member of a clinical or general population, averaging a correct classification rate of 84% (Kraus et al., 2005).

The TOP and real-time client reporting are provided as a free public service. The platform that collects the data and provides the feedback reports is accessible on the Web. The clinician needs to initially register a client in the system. Once the client completes the measure, reports are instantly available for clinician review. Clinicians can arrange for the TOP to be sent to clients directly via e-mail, so that they can complete it at their convenience. The system can also be adjusted to send clients automatic reminders to complete the assessment at a clinician-specified frequency. If the clinician or the clinic decides to administer the TOP at the office, then additional time may be required, as someone must be available to administer the TOP, either in pencil or computer format. BHL's patent-pending technology keeps the client's e-mail information encrypted and separate from their clinical data to ensure confidentiality. If an independent practice or agency would like to share their deidentified assessment data with a larger practice network, BHL has ensured that its data storage methods exceed Health Insurance Portability and Accountability Act (HIPAA) standards. However, if a clinician believes that an IRB should be involved, they can follow their local IRB's guidelines.

The TOP requires very little, if any, training in order to be used properly. The measure itself and the report are user-friendly and easily interpretable. The BHL Web site provides manuals and videos that can be downloaded and viewed, which can assist with the use of the measure or the interpretation of the report.

Using the TOP in Clinical Practice and Naturalistic Research: When and How?

The recommended frequency of administration of the TOP varies from session-to-session to monthly administrations, so that the results can be used and incorporated into treatment planning and treatment plan reviews. The flexibility allows each clinician to

Table 1
Statistics Related to the TOP

TOP subscale	One week test–re-test reliability[a]	Treatment effect sizes (Cohen's *d*)[b]	Reliable change[c]
Work functioning	.90	.44	.62
Sexual functioning	.92	.27	.55
Social conflict	.93	.48	.52
Depression	.93	.91	.52
Panic	.88	.42	.68
Psychosis	.87	.43	.71
Suicidal ideation	.90	.64	.62
Violence	.88	.31	.68
Mania	.76	N/A	.96
Sleep	.94	.57	.48
Substance abuse	.89	.47	.65
Quality of life	.93	.68	.52

Note. Cohen's *d* was calculated using the formula $d = (M1-M2)/\text{SDpre}$, where SDpre referred to the session 1/baseline standard deviation, and was calculated using the formula SDpre = sqrt (sigma[X−M]sqr/*N*).
[a] As reported in Table 7 of Kraus et al., 2005. [b] As reported in Table 6 of Kraus et al., 2011. [c] Calculated using Jacobson and Truax, 1991.

use the measure as frequently as deemed necessary for their particular practice and needs. For example, the TOP can be administered pretreatment to aid case formulation. Given that the initial assessment report can be reviewed with the client in real time, it can be used to start the conversation about the focus of treatment. While gathering information from the client during the session, therapists can also use the TOP report to provide additional insights into the client's level of distress across various areas of symptoms and functioning. Additionally, the TOP can be administered at both pre- and posttreatment as a measure of change that can provide specific information about issues that have improved, those that have not, and dimensions of functioning that may have deteriorated. However, a more frequent session-by-session application of an outcome measure allows clinicians to track the client's progress in treatment, as well as determining whether the treatment primary goals are being achieved.

For example, at the Pennsylvania Practice Research Network (PRN), a network composed of clinicians in independent practice (see Castonguay, 2011), the clinicians administer the TOP before treatment starts as part of routine clinical practice. Additionally, these clinicians have designed a research protocol using the TOP in which the measure is administered at session seven and posttreatment. The principal reason that led to the decision to administer the TOP at session seven was that in a previous study conducted by these clinicians (Castonguay, Boswell, Zack, et al., 2010), the average number of sessions in independent practice was found to be seven. Thus, by having clients complete the TOP at session seven, clinicians were able to have an assessment that could both measure treatment progress and outcomes.

In another PRN, the Penn State University clinic (Castonguay et al., 2004), the therapists administer the TOP at every session. Such frequency is possible because the TOP is a short measure and only takes a few minutes to complete. As part of their clinical routine, clients are asked to complete the TOP before their therapy appointments in the waiting room using netbooks, which send the clients' responses to BHL electronically. The TOP is then scored by BHL, and a report is returned to the therapist (typically within a minute after client completion). The report, which graphically displays how the client deviates from a nonclinical population for each of the 12 domains measured by the TOP, can be reviewed before the first session—if the therapist so desires. This allows the therapist to have access to the client's symptomatology and level of functioning, which can be contrasted with in-session diagnostic assessments. Ongoing assessments can provide useful information with graphed contrasts (again for each of the dimensions measured) of the client's prior administrations. In this way, change (positive and negative), or lack of change, can be monitored session-by-session.

As an example, a therapist treating a rather treatment-resistant obsessive compulsive disorder (OCD) client used the ongoing assessment of the client's TOP scores. He tracked the client's anxiety (via the panic scale) each week with two goals in mind: first, he monitored the client's adherence to the exposure homework assignments by noting the presence of elevated anxiety scores; second, as treatment progressed, he used the consistently decreasing anxiety scores to illustrate to the client that positive change was taking place as demonstrated by the contrasting graphs. As the outcomes were based on the patient's self-report data, the information helped to explore the patient's in-session verbalizations that treatment was not working. The result was greater patient engagement and the facilitation of a stronger therapeutic relationship.

As mentioned earlier, the TOP scores are processed within several minutes, allowing therapists to choose whether to review the results before the patient sits down for their first interview. If they make that personal choice, this can help prepare them for certain topics that the clients may want to bring up during the session or may direct them to areas that should be discussed (such as suicidality). The TOP also helps clinicians anticipate a client's current state of mind, including whether they typically minimize or exaggerate their symptoms, or whether emotions like shame may prevent them from fully discussing issues that may include concealed hostility, violence, or substance abuse issues. For example, a therapist used the TOP's Substance Abuse scale to confront a client over her problematic alcohol use. During the initial meetings, the client denied any problems with her own drinking, and instead, repeatedly reported that her boyfriend and mother were the ones that experienced problems with their alcohol intake. The therapist approached the subject by showing her the TOP scores, explaining that her scores were five standard deviations above the healthy population in the Substance Abuse scale. This helped the patient to see that, from her self-report, alcohol use was a real problem, one that became an important aspect of treatment.

In another example, a clinician used initial TOP suicide scores to uncover a high degree of suicidal ideation that was actively denied during the first interview. The patient not only denied any suicidal ideation or plan, but when shown the individual items of the TOP also dismissed his answers stating that he must have misread the questions. However, when a month later, the suicide score was similarly elevated and the scores and individual items were shown again to the patient, the client finally pulled back his long sleeves that hid wounds from a previous suicide attempt and acknowledged that (i) the same feelings had returned and (ii) that the "chief complaint" for coming to treatment was not really the social issues that had been the focus of the first month of treatment but rather the suicidal ideation. He explained that both shame and fear of rejection were the basis for the slow pace of the disclosure. He also expressed that the TOP items provided a way for him to "test" his therapist to see whether she would neither let the issue go and at the same time, show no sign of rejection, as she became increasingly suspicious of his denials. "Obviously, I wanted you to know," he told his therapist toward the end of treatment. "I just needed to see how you were going to handle it . . . . If I wanted to keep it away from you, I would have denied it on the TOP questions."

As will be discussed later, this example illustrates how a well-chosen outcome tool can provide another, maybe less threatening, channel for communication and an alternative forum for disclosure of highly sensitive and personal information. How the therapist responds to these "partial disclosures" is essential to fostering a good therapeutic relationship. Rather than reacting to these partial disclosures with displeasure (directed either toward the client or the questionnaire as if one form of communication is more valid or truthful), the "news of a difference" might be best celebrated with an inquisitive eye toward the whys—"why did the client answer the question this way when I asked them about this directly as compared with their answers on the questionnaire; and what helped them feel safer or more comfortable telling me about this?"

Yet, it is important to recognize that partial disclosures need to be addressed with tack and good clinical timing. As an example of how not to respond to the partial disclosures on outcome tools, a therapist called the second author and primary developer of the TOP to complain about how the TOP's outcome report had caused an early treatment termination. He explained that a middle-aged woman came to treatment to discuss issues that she was having with her boss at work. On the initial TOP evaluation, the report came back with a number of diagnoses to consider including a sexual pain disorder related to her endorsement of pain during intercourse. With the report scored and returned between the first and second session, the therapists walked into the next session with one question: "Why didn't you tell me about this in our first meeting?"

We agree that this is the right question for the therapist to be asking, but not to harangue the patient with it. No matter how kindly or softly one asks this direct question, it comes with the implicit message that the patient has done something wrong and "should have" made the disclosure earlier. According to the therapist, the patient rose from her chair and walked out, never to return to treatment. If faced with a similar situation, we suggest that a less confrontational approach might be to say something like: "Here's a copy of the report that comes back from the TOP questionnaire you took last week. Let's go over it together. In addition to the things we talked about last week, there are a couple other areas for which you have expressed some concern. I can help you with these issues as well when you feel comfortable talking about them. Is it all right if I ask you about those now?"

After training tens of thousands of clinicians on how to use outcome tools, we have learned that the primary fear that clinicians have is that patients will minimize their problems or pathology on outcome tools. Obviously this can happen, but what is far more common is having patients, like those described earlier, reveal more pathology on a questionnaire than they feel comfortable doing in initial interviews. Recognizing that great sensitivity must be used when addressing these disclosures can go a long way in helping to use these disclosures productively.

What are the Benefits of Using the TOP?

Repeated TOP administrations have been found to be helpful for clients. Some clients have reported that completing the TOP regularly has been informative for them in terms of their own understanding of their difficulties and clinical progress. Clients usually have a hard time remembering how they felt more than 2 weeks ago. Thus, having an assessment such as the TOP provides comparison points and even shows subtle changes in symptomatology (like suicidal ideation) even before it becomes a conscious issue. This sensitivity to detect slight improvements often helps clients see that they are making recognizable progress. As a result, this can improve motivation, hope, and may further cement a strong therapeutic alliance.

The TOP can also provide information that is not disclosed in the initial interview. As previously mentioned, these problems are often shame-based issues that can be difficult for many people to talk about when making eye contact with a relative stranger—regardless of the alleged safety afforded them by the psychotherapy privilege. An example of such an instance involved a client who had come in for therapy for a panic disorder. The client's presenting concerns revolved around his numerous panic attacks and avoidant behavior. However, when the treating clinician looked at the TOP that had been completed after the first session, the Substance abuse scale was significantly elevated. After initial discussion with the client, the clinician believed that the elevation might have been related to the client's age-group (since he was a college student) and the cultural norm at the university he attended. Based on this judgment, the clinician only focused the next sessions on decreasing the panic attacks and panic disorder symptoms. At session seven, the TOP scores confirmed the client's self-report that the panic disorder symptoms had decreased significantly, but they also continued to show elevated substance abuse scores, which were subsequently accompanied by a driving-under-the-influence arrest. This was an important lesson for the therapist who concluded that he should have addressed the elevated TOP substance abuse scale at the beginning of treatment, even if such issues had not come up in the initial interview.

The ability of outcome tools to enhance patient–therapist communication is not limited to individual treatment. In couples therapy, one clinician used TOP reports from both the husband and wife to begin a mutual dialogue of how to integrate the individual issues that were assessed on the TOP into the relational work. Both husband and wife agreed to complete the TOP and share the results with each other in treatment. In reviewing the wife's report, the husband and therapist had questions related to the wife, listing a recent trauma that was causing a high level of stress on Axis IV. According to the therapist, the wife took a deep breath and started talking about a recent rape that she had been unable to disclose to her husband. When asked whether the questionnaire had helped her "break the ice," the wife agreed.

As many of the aforementioned examples illustrate, the TOP can be viewed and used as a communication tool. It allows the client to stick a toe in the water and see how therapy is different than other relationships and how this specific therapist will respond. Clients may have thoughts or symptoms that are not easily shared with a therapist at intake. However, answers to these questions are often captured by the TOP. As another example, a private practitioner used the TOP results to learn about a client's suicidal ideation, ideation that had not been discussed during the session. This realization led the clinician to follow-up with the client postsession, establish a safety contract, and discuss the issues in treatment.[1] This situation exemplifies how outcome data can provide a "corrective experience" for skeptical therapists who are worried that data collection interferes with, rather than facilitates and augments, the process of therapy and change.

Designed by clinicians, the TOP was constructed to provide useful clinical information that would not be available otherwise. Among its core features, the TOP includes benchmarks compared with the general population, as well as to clinical populations. Each of these reference groups provides different kinds of alerts and information to providers, such as whether the patient is potentially off-track or whether the treatment is potentially causing harm. In addition, TOP domains have been linked to evidence-based practices (such as empirically based principles of change

[1] Provider Focus. Blue Cross Blue Shield of Massachusetts Newsletter, 1-12. https://www.bluecrossma.com/staticcontent/newsissues/Provider%20Focus-April%202008-47.pdf

that cut across different forms of treatments [Castonguay & Beutler, 2005]), allowing BHL to provide clinicians with an extensive source of information regarding the treatment of specific clinical problems. The TOP also alerts the clinician to the possibility that a patient may be at high risk for hospitalization within the next 6 months. These kinds of alerts were included to assist providers in independent practice in formally evaluating risk, as well as to provide the basis of more collaborative decision making between insurance companies and treating providers.

The Raison d′être of the TOP: Documenting Change

Underlying the benefits described previously is the primary mission of the TOP: documenting change, both positive and negative. Research has demonstrated that 92% of clients show clinically and statistically significant change on at least one TOP dimension as they continue to receive treatment (Kraus et al., 2005). More recent findings (Kraus et al., 2011) also reveal that while few, if any, therapists demonstrate proven effectiveness on all of the TOP's multiple areas of assessment, the vast majority of therapists seem to be particularly effective in treating certain types of clients. Only 4% of clinicians have been shown to be ineffective in treating all types of client problems as measured by TOP. The other 96% of clinicians are reliably effective at treating at least one diagnostic or functional cluster (e.g., depression, substance abuse, mania). In fact, the average clinician appears very effective at treating five of these TOP measure clusters. Interestingly, however, showing superior effectiveness in treating a number of specific clinical problems does not preclude therapists to have difficulties in helping clients who are experiencing other particular concerns. In the 2011 study, the therapist who was measured as the best at treating depression was among the worst at treating mania. In fact, he or she was one of the two clinicians in the study that was reliably harmful in treating manic symptoms, with the average client showing statistically and clinically significant worsening of symptoms. Perhaps not surprisingly for experienced clinicians and supervisors, all of the nearly 700 clinicians in the study had areas of strengths and weaknesses. As clinicians, we may have a responsibility to identify these areas of weakness and either find the supervision or training necessary to improve our effectiveness or refer new patients who may need care in these domains to other therapists who are effective.

This discussion illustrates the importance of having a dimensional assessment of psychopathology. Because the TOP is comprised of 12 subscales, it is able to discriminate between the aspects of the client's functioning for which therapy may have more or less impact. This can be crucial for guiding the therapist's decisions, not only in terms of what should be the focus of treatment, but also when to terminate, or titrate, therapy. Directly related to this important point, a study on CBT for GAD (a treatment that focused primarily on reducing worry and physiological symptoms) found that the level of interpersonal problems at the end of therapy was predictive of relapse (Borkovec et al., 2002). This suggests that although therapists and clients might consider terminating a treatment when the presenting symptoms have decreased, this decision might be premature unless it is based on a reliable assessment of other conditions that can affect the maintenance of change.

In addition, research findings obtained with the TOP have also shown that 58% of clients experience clinical deterioration on at least one clinical dimension during treatment (Kraus et al., 2005). Although many clinicians may see the merits (in terms of simplicity and efficiency) of relying on a single measurement of pathology (whether it measures a specific dimension of functioning or a global level of distress), these negative changes may not have been detected with a one-dimensional tool. An overly narrow assessment would have missed potentially unintended negative consequences of therapy and a global score would have likely combined both improvement and worsening of symptoms, which then might be interpreted as showing no change. Therefore, an outcome tool that addresses the multiple directions and areas in which someone can change, such as the TOP, may be a useful instrument in augmenting the positive changes and help ameliorate the negative ones. Such multifaceted feedback may help enrich and modify the case formulations as treatment begins and therapy progresses. Needless to say, the use of an instrument that can provide feedback to early career therapists (and their supervisors) about aspects of functioning with which they may (or may not) be particularly effective in treating, as well as other areas with which they may cause harm, is likely to be a very relevant tool for training (Castonguay, Boswell, Constantino, et al., 2010).

Challenges in Using the TOP in Clinical Practice

Many clinicians start using outcome assessments with the fear that clients will be resistant and opposed to formalized assessments. Many similarly fear that the process might interfere with forming a therapeutic alliance or the flow of the therapy hour. It has also been our experience that some clinicians fear that an outcome assessment will show evidence of their ineffectiveness. However, after program implementation, most clinician and patient experiences are quite different. For example, Blue Cross and Blue Shield of Massachusetts conducted clinician focus groups and competitive requests for proposals in selecting the TOP. The program carefully tracked patient and clinician feedback. After the first 6 months of the program with 40,000 patient administrations, Blue Cross concluded:

> Member acceptance has been quite high with a very small number of complaints; and, after a period of challenging exchanges with providers and their associations, it has become widely accepted that the measurement of outcomes in behavioral health is positive.[2]

As we previously mentioned, most clinicians appear to be effective in addressing multiple types of psychopathology. We would venture to guess, however, that for the most part, clinicians would not be able to reliably predict which aspects of a client's life they are particularly effective at treating. Rather than confirming unfounded fears, we believe that the use of the TOP can lead to a validation of the diversity of the therapist's strengths—some of which are unanticipated or unappreciated by clinicians themselves. Put in other words, the TOP can facilitate for the therapist what most of them try to foster for their clients—a corrective experience (Castonguay & Hill, in press).

[2] Blue Cross and Blue Shield of Massachusetts (2008) Behavioral Health Outcomes Challenges and Opportunities. Unpublished report.

Another concern regarding the implementation of the TOP is a pragmatic one, such as time and length of the administration. To address these concerns and mitigate the administrative burden, the TOP allows flexibility so that it can adapt to different clinical practices. As discussed above, once providers register their patients with their e-mail addresses on the BHL Web system, BHL will then take on the task of sending the TOP to the clients, at the frequency requested, with daily reminders until the assessment is complete. Clients have the option of changing the frequency with which they receive the TOP, as well as withdrawing from the service. Furthermore, if desired, clients can be contacted and asked to complete the TOP even after treatment as a means to continue to monitor their symptoms, progress, and help prevent relapses.

With BHL's online system, clients have the option of completing the TOP online anytime and anywhere before their scheduled therapy assessment. Independent of whether the TOP was completed online, or via paper and later faxed to BHL, clinicians have the option of receiving the report via fax, as a secure email attachment, or by logging onto WellnessCheck.net and downloading the report.

Therapist confidentiality concerns may pose an additional challenge in implementing the TOP. Clinicians might fear that the TOP data could potentially be used by insurance companies and/or employers in punitive ways. Recently, there has been an increasing wish from employers and state and federal governments to use quality data when making purchasing decisions. This trend and driving force cannot be ignored. However, a possible way of achieving a balance between an employer's needs and clinicians' concern is to encourage data use and collection for 6 months to a year before the data is used to profile clinicians. This time gap is likely to allow clinicians to become familiarized with outcome measures and modify practice patterns to maximize outcomes through quality improvement and continued education.

Drawbacks of Standardized Assessments

Caution should be applied when making use of standardized ongoing assessments. The main drawback is the risk of using the results of such measures without contextualizing them. Such misuse could take place, for example, if the decision to discharge a patient from an inpatient unit fails to take into account that the patient may be exaggerating or minimizing their symptoms in order to achieve a desired goal. Another example would be to interpret a slow rate of change (when working with a traumatized client for instance) as evidence for the need to hasten or intensify treatment before considering issues such as the client's readiness to change, quality of the therapeutic alliance, and/or the therapist's own competence in implementing treatment (with this particular client or a clinical disorder in general). Along the same line, another misuse of standardized assessments would be for an insurance plan to use the information to manage benefits such as the number of sessions people should be allocated, based on these reports.

A particularly important point to consider when using outcome measures is the fact that after the first intake assessment, the scores may actually worsen. This may be due to various reasons: clients may attempt to normalize their problems or may not be comfortable disclosing the extent of their distress at intake. After the establishment of a safe therapeutic relationship, however, they may become either more aware of their distress or feel safer to disclose their true suffering and impairment.

These examples illustrate the importance of taking into account that numerous factors may affect honest assessments, as well as the need for frequent evaluation of the client's symptoms and level of functioning. At a conceptual level, it is important to recognize that standardized instruments, such as the TOP, provide one operationalization of the constructs of psychopathology and mental health. Furthermore, the TOP is a self-report measure, which makes it susceptible to some of the limitations associated with self-report assessments. For example, clients may inadvertently distort their answers due to social desirability, biases in how they recall their symptoms, and poor insight into their distress. Thus, the TOP and other standardized assessments should be used as one tool out of many sources of information that will aid in making a complete case formulation.

Helping Pave a Two-Way Street

As we previously mentioned, although the raison d′être of the TOP is to document change, this mission has, from its beginning, taken place within the hope of building a stronger bridge between practice and research. We conclude this article by briefly describing three ways that the TOP can play a role in paving an exciting two-way street between scientific advance and clinical improvement: (1) providing an empirically valid instrument to help with important clinical tasks; (2) offering an infrastructure for groups of clinicians to conduct studies in naturalistic setting, and (3) fostering a seamless integration of clinical and empirical tasks.

The TOP is a measure that was developed by clinicians to be used primarily in clinical practice. A short, valid, and informative measure like the TOP can help practitioners with many crucial clinical tasks. It can help clinicians in quickly formulating both a specific and comprehensive treatment plan, not only by providing a reliable assessment of many of the complaints that clients present to therapy (e.g., depression, anxiety, substance abuse), but also of several factors that are frequently associated with these complaints. In a recent overview of basic research findings in psychopathology, Castonguay and Oltsmann (in press) concluded that for the most common forms of disorders seen in independent practice, comorbidity is the rule rather than the exception, and that the majority of disorders are associated with suicidality, sleep difficulties, sexual dysfunction, health concerns, as well as interpersonal and occupational problems. The clinical implication of these findings is that therapists will benefit greatly by using an instrument such as the TOP to measure an extensive variety of symptoms and clinical factors at the beginning, during, and before terminating therapy.

Ongoing assessments can inform other clinical areas, such as client referrals. The strategies on which referrals are typically based, include the following: (a) convenience, such as an opening in a therapist's caseload, or insurance coverage, (b) personal criteria, such as intuition that a certain client would relate well with a certain therapist, or having a relationship with a therapist (while not knowing what the therapist is like in-session), and (c) "expertise," determined either by word of mouth, knowing that a certain clinician treats a high number of clients with a specific disorder, or based on their graduate school training, or having attended workshops on a specific topic. However, the evidence that such strat-

egies lead to good client–therapy matching and, ultimately, good treatment outcome, are not known. Moreover, some research findings can legitimately put into question the "predictive validity" of these decisions. For instance, research suggests that the therapist level of experience is a weak contributor to treatment outcome (Beutler et al., 2004). In addition, therapists' ability to choose clinicians who are experts at certain treatments and/or disorders has not been found to consistently yield positive outcomes, and instead, experts have shown to have wide variability in terms of outcome, and even sometimes provide harmful interventions (Luborsky et al., 1997).

The empirically based information provided by the TOP may increase the supply of well-matched referrals. This could be done by having therapists share their risk-adjusted benchmarks that highlight relative strengths and weaknesses with respect to each of the 12 dimensions of symptoms and functioning it measures. As we mentioned earlier, the results of a recent study (Kraus et al., 2011) indicated that 96% of clinicians excelled at the treatment of at least one dimension measured by the TOP. Clinicians who provide treatment in their area of expertise seem to have triple the treatment effect sizes compared with the average clinician. The same study also demonstrated that a large percentage of clinicians provide treatment in areas in which they may have minimal impact and/or may potentially cause unwanted harm to patients. These findings raise the interesting possibility that, by providing referral sources with objective data on strengths, the number of well-matched referrals could increase. The result for the clinician might be an increased sense of professional accomplishment and improved client outcomes, and for the patient, an increased chance of successful treatment.

BHL has patented the use of outcome data in making referral recommendations and has set up a Web service—OutcomeReferrals.com—to facilitate better therapist–patient matching. Based on Kraus et al. (2011) data, this may lead to a substantial increase of the effect size of naturalistic treatments. Using this service, clinicians can screen prospective patients for a good fit. By focusing their practice on clients with which they are consistent and effective (and, if they so desired, by seeking training or supervision on clinical problems they appear to be less effective), they may well document superior overall outcomes in their clinical practice.

The TOP can also contribute to the viability of the scientific-practitioner model by facilitating clinicians' research within their own work setting. The following example describes how one of BHL's clients received the Joint Commission's highest Codman Award for quality improvement. This Texas-based agency used TOP monthly aggregate reports that are automatically generated by BHL for each clinician, service (e.g., a partial hospitalization program), and agency. These reports highlight areas of strength and weakness and are benchmarked, risk adjusted, and mapped against similar professionals across the world (or local regions) based on a growing database of over a million patients (Kraus & Castonguay, 2010). Using these tools, this adolescent residential treatment program discovered that their services were comparatively poor in treating adolescent anger and violence issues. This led to an integration of a systematic, focused, and empirically based training of the clinical staff within the day-to-day functioning of the agency. The repeated assessment of the TOP (also part of their clinical routine) was able to document the quality improvement achieved in the specific area of functioning targeted by the treatment (Adelman, 2005, 2006, 2007, 2008). To further improve the ability of providing the most effective treatment to each of their clients, this clinical program (in collaboration with researchers at Penn State) is now investigating whether client past and interpersonal problems, as measured at pretreatment, can predict outcome and the quality of the alliance.

The TOP has also played a crucial role in PRNs, which are aimed at building scientifically rigorous and clinically relevant studies based on the full collaboration of researchers and clinicians in all aspects of research (from the generation of ideas to be investigated, the design and implementation of the study protocol, to the dissemination of the findings). At this point in time, the TOP has served as the main outcome instrument in the assessment battery of two PRNs (one within the context of independent practice and the other within a psychological training clinic) that have been described elsewhere (see Castonguay, 2011). Interestingly, although the TOP (or any outcome measure) may at first appear, for some clinicians, as a research tool required for their engagement in a PRN, it can also become one of the main benefits of such professional commitment. For example, a practitioner who has been part of the Pennsylvania Psychological Association PRN reported having been initially resistant to using the TOP because she already used other measures in her clinical routine. However, after using the TOP in the context of a research protocol that she helped design, she reported the TOP to be very informative and efficient and now uses it in her standard practice.

Whether or not a clinician is a member of a PRN, he or she can use the TOP to conduct research within his or her own clinical practice (using single-case methodology, for instance). Because BHL aggregates data for each individual therapist, a clinician can easily assess change (at any kind of frequency) for particular types of clients, for specific types of interventions, or for any kind of interaction between client characteristics and treatment variables that may be of interest. The use of the TOP as part of research conducted within natural practice provides clinicians (as a single practitioner or as part of a group) information about the process and impact of their intervention, and it can also facilitate the seamless integration of science and practice. As described elsewhere (Castonguay, 2011), it may well be that a particularly fruitful strategy to foster the growth of the boulder model is to create conditions under which research and clinical tasks are intrinsically and completely confounded, intertwined, and indistinguishable. Specifically,

> "It could be argued that clinicians truly integrate science and practice every time they perform a task in their clinical practices and are not able to provide an unambiguous answer to questions such as: "Right now, am I gathering clinical information or am I collecting data?" or, "At this moment, am I trying to apply a helpful intervention with my client or am I implementing a research task?" Frequently, setting up rigorous empirical investigations that will lead them to answer these questions by saying, "Perhaps both," may be the most fruitful and exciting pathway to bridge research and practice" (pp. 352–353).

Whether it is simply to observe change of particular types of (or all of his or her) clients or to investigate the impact of particular interventions, the clinician's use of the TOP (and any other standardized outcome measure) within his or her own practice can contribute to the actualization of the scientific-practitioner model. He or she may do so by collecting data that could simultaneously

serve clinical and empirical purpose or, using different words, by conducting clinically syntonic research (Castonguay, 2011).

References

Adelman, R. (2005). Reducing adolescent clients' anger in a residential substance abuse treatment facility. *Journal on Quality and Patient Safety, 31,* 325–327.

Adelman, R. (2006). The Angry Adolescent & Constructivist REBT. In P. Cummins (Ed.), *Working with anger: A practical constructivist approach*. London: John Wiley & Sons.

Adelman, R. (2008). Methods of reconstruction with adolescent substance abusers: Combining REBT and constructivism. In J. D. Raskin & S. Bridges (Ed.), *Studies in meaning 3: Constructivist psychotherapy in the real world*. New York: Pace University Press.

Adelman, R. (2007). *Reducing anger in adolescents: An REBT approach*. Center City, MN: Hazelden Press.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed.; pp. 227–306). New York: Wiley.

Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70,* 288–298. doi:10.1037/0022-006X.70.2.288

Castonguay, L. G., & Beutler, L. E. (Eds.). (2005). *Principles of therapeutic change that work.* New York: Oxford University Press.

Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist, 65,* 34–49. doi:10.1037/a0017330

Castonguay, L. G., Boswell, J. F., Zack, S., Baker, S., Boutselis, M., Chiswick, N., . . . Grosse Holtforth., M. (2010). Helpful and hindering events in psychotherapy: A practice research network study. *Psychotherapy: Theory, Research, Practice, and Training, 47,* 327–344. doi:10.1037/a0021164

Castonguay, L. G., & Hill, C. E. (Eds.). (in press). Transformation in psychotherapy: Corrective Experiences across cognitive behavioral, humanistic, and psychodynamic approaches. American Psychological Association Press.

Castonguay, L. G., & Oltmanns, T. G. (in press). Psychopathology research and clinical interventions: Broad conclusions and general recommendations. To appear in Psychopathology: Bridging the gap between basic empirical findings and clinical practice. New York: Guilford Press.

Castonguay, L. G., Pincus, A. L., Arnett, P. A., Roper, G., Rabian, R., & Borkovec, T. B. (2004, November). Psychology training clinic as a research Practice network: Integrating research and clinical practice in graduate school. Paper presented at the annual meeting of the North American Society for Psychotherapy Research. Springdale, AZ.

Castonguay, L. G. (2011). Psychotherapy, psychopathology, research and practice: Pathways of connections and integration. *Psychotherapy Research, 21,* 125–140. doi:10.1080/10503307.2011.563250

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19. doi:10.1037/0022-006X.59.1.12

Kraus, D., & Castonguay, L. G., (2010). Treatment Outcome Package (TOP): Development and use in naturalistic settings. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds), *Developing and delivering practice-based evidence: A guide for the psychological therapies.* New York: Wiley. doi:10.1002/9780470687994.ch7

Kraus, D. R., Castonguay, L. G., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research, 21,* 267–276. doi:10.1080/10503307.2011.563249

Kraus, D. R., Seligman, D., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The treatment outcome package. *Journal of Clinical Psychology, 61,* 285–314. doi:10.1002/jclp.20084

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice, 4,* 53–65. doi:10.1111/j.1468-2850.1997.tb00099.x

Received February 20, 2012
Accepted February 23, 2012 ■